k-匿名下通过本地差分隐私实现位置隐私保护 *

刘振鹏 [a,b]，苗德威 [a]，刘倩楠 [a]，李瑞林 [a]，李小菲 [b†]
(河北大学 a. 网络空间安全与计算机学院; b. 信息技术中心, 河北 保定 071002)

摘　要：针对用户位置隐私保护过程中攻击者利用背景知识等信息发起攻击的问题，提出一种面向移动终端的位置隐私保护方法。该方案通过利用 k-匿名和本地差分隐私技术进行用户位置保护，保证隐私和效用的权衡。结合背景知识构造匿名集，通过改进的 Hilbert 曲线对 k-匿名集进行分割，使用本地差分隐私算法 RAPPOR 扰动划分后的位置集，最后将生成的位置集发送给位置服务提供商获取服务。在真实数据集上与已有的方案从用户位置保护、位置可用性和时间开销方面进行对比，实验结果显示，所提方案在确保 LBS 服务质量的同时，也增强了位置隐私保护的程度。
关键词：RAPPOR；k-匿名；Hilbert 曲线；位置保护
中图分类号：TP309.2　　doi: 10.19734/j.issn.1001-3695.2021.12.0698

Location privacy protection through local differential privacy under k-anonymity

Liu Zhenpeng[a, b], Miao Dewei[a], Liu Qiannan[a], Li Ruilin[a], Li Xiaofei[b†]
(*a. School of Cyber Security & Computer*, *b. Information Technology Center*, *Hebei University*, *Baoding Hebei* 071002, *China*)

Abstract: Aiming at the problem that attackers use background knowledge and other information to launch attacks in the process of protecting user location privacy, this paper proposes a location privacy protection method for mobile terminals. This solution uses k-anonymity and local differential privacy technology to protect the user's location to ensure the trade-off between privacy and utility. This scheme combines background knowledge to construct anonymity sets, uses the improved Hilbert curve to segment the k anonymous set, uses local differential privacy algorithm RAPPOR to perturb the divided location sets, and finally sends the generated location sets to location service providers to obtain services. Comparing this scheme with existing real data set schemes in terms of user location protection, location availability and time overhead, the experimental results show that the proposed scheme not only ensures the quality of LBS service, but also enhances the degree of location privacy protection.
Key words: RAPPOR; k-anonymity; Hilbert curve; location protection

0　引言

互联网技术、卫星定位技术和移动设备高速发展，基于地理位置的服务(location-based service ,LBS)受到广泛应用[1,2]。用户在享受位置服务带来的便捷的同时，LBS 服务所带来的用户位置隐私泄露问题也造成了极大的困扰。恶意位置服务提供商(location service provider,LSP)通过用户位置获取用户的敏感信息，严重侵害了用户的隐私。因此，位置隐私保护是用户隐私保护研究中的热点问题 [3]。

在位置隐私保护研究中，基于匿名的 K-匿名技术被广泛应用，该技术最早由 Sweeney[4]提出，其核心思想为使用属性泛化使单个数据与其他 k-1 个数据无法区分。Gruteser M 等人[5]首次将 k-匿名技术作为位置隐私保护手段，通过四叉树搜索构造 k-匿名位置模型，保证匿名区域不小于一定值。但该方法增加了时间开销，容易造成匿名位置过剩，且匿名的 k 值是一样的，无法满足用户的个性化选择。为解决生成匿名位置过剩问题，Kido 等人[6]使用随机策略生成 k-匿名集，降低了通信成本，但其没有考虑一些不合理的虚假位置，降低了安全性和服务质量。Zhu 等人[7]在 Kido 的基础上添加位置缓存机制，设计了 MobileCache 系统，通过减少查询次数降低了资源开销，提高了资源利用率。叶阿勇等人[8]通过考虑服务相似性生成匿名区，提升了服务质量，但他们都没有考虑背景知识对隐私保护的影响。Yin 等人[9]将 k-匿名方法与假名方法相结合，通过 k 的最大值和最小值选择相应的匿名方法，改进了 k-匿名位置保护方法。Jin 等人[10]设计了一种基于信任的位置隐藏机制，通过在匿名区域中添加 k-1 个可信用户，保证每个用户都能达到其所需的匿名级别，但该机制需要依赖集中式匿名服务器。Ling 等人[11]为解决不可信匿名服务器的问题，构造了一种基于偏移网格的分布式位置隐私保护机制。通过历史查询概率将位置区域划分为位置网格，选择 k-1 个网格坐标构成匿名集，使得不可信匿名服务器难以获取用户真实信息。Zhang 等人[12]结合 k-匿名思想，利用不规则多边形生成算法，生成多边形匿名区域。通过设置密度参数实现空间匿名性，构造虚拟位置。闫光辉等人[13]通过服务相似性构造相似地图，从相似地图中选取与用户真实位置查询结果相似的兴趣点，并结合背景知识生成熵最大的匿名集，随机选取匿名集中的一个位置来完成查询服务，在保证用户隐私安全的同时尽可能提高服务质量。杨洋等人[14]设计了一种基于历史查询概率的 k-匿名集选取算法，从地理分布和零查询两个方面提升了位置隐私的安全性，但其在构建

收稿日期：2021-12-13；**修回日期：**2022-03-01　**基金项目：**河北省自然科学基金资助项目(F2019201427)；教育部“云数融合科教创新”基金资助项目(2017A20004)
作者简介：刘振鹏(1966-)，男，河北保定人，教授，博导，博士，主要研究方向为网络信息安全、隐私保护等；苗德威(1998-)，男，河北邯郸人，硕士研究生，主要研究方向为信息安全、位置隐私保护；刘倩楠(1999-)，女，陕西渭南人，硕士研究生，主要研究方向为区块链、隐私保护；李瑞林(1996-)，男，河南周口人，硕士研究生，主要研究方向为隐私保护、入侵检测；李小菲(1979-)，女(通信作者)，河北保定人，工程师，硕士，主要研究方向为网络信息安全、隐私保护(lixiaofei@hbu.edu.cn).

匿名集时需要进行离散处理，增加了匿名集的生成时间。

差分隐私(differential privacy, DP)最早由 Dwork[15]于 2006 年提出，通过严谨的数学证明，可以保证用户隐私不受攻击者所知的背景知识攻击以及某个数据变化的影响。袁健等人[16]设计了一种拉普拉斯机制和匿名组相结合的 LBS 轨迹保护算法，对 LBS 用户的真实位置进行多轮加噪生成匿名组，用匿名组来获取 LBS 服务，解决了差分隐私实现轨迹隐私保护时产生的隐私预算过度依赖问题，增强了轨迹隐私保护效果。Wang Jie 等人[17]提出了一种基于差分隐私扰动的位置保护方法，利用 Hilbert 曲线将位置映射到一维空间中，通过 Laplace 噪声对位置信息进行扰动，将扰动后的位置信息发送给服务商来实现位置保护。Zhang 等人[18]采用基于最大-最小距离的多中心聚类算法，生成多组候选虚拟对象，选择最优虚拟候选集实现 k-匿名。Zhang 等人[19]提出了基于差分隐私的位置隐私保护方案，该方案包括均值算法和匿名算法，通过 LapLace 机制保护用户的位置隐私，利用指数机制保护用户的查询隐私。

通过差分隐私保护敏感信息需要依赖可信第三方(fully-trusted third party ,TTP)数据收集器，但在真实环境中第三方的安全性往往不能得到保证。因此，有学者提出了本地差分隐私(local differential privacy ,LDP)概念[20~22]，用户可以在本地对敏感数据进行处理，从而避免不可信第三方的泄露问题。Wang 等人[23]提出了一种基于 LDP 的位置连续上传保护方案，使用 Hilbert 曲线根据区域内用户位置数量动态划分子区域，通过本地差分隐私对位置进行扰动，将扰动后的位置上传到服务器，但其数据可用性降低。Wang 等人[24]让参与者根据当前位置的个人隐私需求，选择两种不同的本地差分隐私扰动方法，RAPPOR 和 k-RR。对参与者的位置进行区域分割，用选择的扰动方法对位置区域进行扰动，将扰动后的位置发送到数据收集服务器用于数据分析。

针对上述方法中的问题，本文结合 k-匿名和本地差分隐私技术，提出一种无须 TTP 且能够抵御背景知识攻击的本地差分隐私扰动方案，在保证性能的同时降低了攻击者获取用户信息的概率，进一步提高了用户位置隐私安全性。

1 相关概念

1.1 背景知识

背景知识指用户在特定位置发送位置服务请求的概率。由于实际生活中，人们在不同位置获取 LBS 服务的概率是不同的，因此攻击者可能通过此类信息推断匿名用户的真实位置等敏感属性。将某个地区划分为 $N\times N$ 个位置区域，每个区域 loc_i 被访问的概率 p_i 为该区域的查询次数 m_i 与整个地区查询次数 M 的比值，记为

$$p_i=\frac{m_i}{M} \tag{1}$$

其中，$i=1, 2, \cdots, N^2$，且 $q_i=\frac{p_i}{\sum_{j=1}^{k}p_j}$。

1.2 位置熵

在不考虑背景知识的前提下，k-匿名保护下用户真实位置被识别的概率为 $q_i=\frac{p_i}{\sum_{j=1}^{k}p_j}$。设 q_i 为 loc_i 是真实位置的概率，则

$$q_i=\frac{p_i}{\sum_{j=1}^{k}p_j} \tag{2}$$

其中 $i=1, 2, \cdots, k$，且 $dis(r_i,Z)=\sqrt{(x_r-x_Z)^2+(y_r-y_Z)^2}$。

位置熵可以用来预估匿名位置集的隐私保护强度。熵值越大，位置集越无序，隐私保护程度就越高。公式如下：

$$dis(r_i,Z)=\sqrt{(x_r-x_Z)^2+(y_r-y_Z)^2} \tag{3}$$

由式(3)可知，当所有 q_i 都为相等的概率值时,位置熵 H 最大，此时 $dis(r_i,Z)=\sqrt{(x_r-x_Z)^2+(y_r-y_Z)^2}$。

1.3 Hilbert 曲线

Hilbert 曲线用作将 s 维空间 Rs 映射到一维空间 R，表示为 H：Rs->R。如果点 $p\in Rs$，然后 $H(p)\in R$；也就是说，$H(p)$ 是 p 对应的 H 值。对于点集{p_1，p_2，$\cdots$，p_n}，H{p_1，p_2，$\cdots$，p_n}={$H(p_1)$，$H(p_2)$，$\cdots$，$H(p_n)$}。Hilbert 曲线的编码规则如图 1 所示。

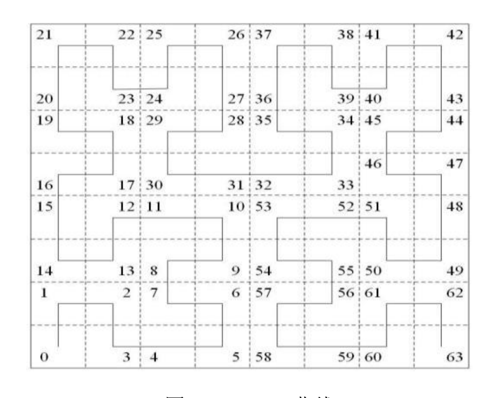


图 1 Hilbert 曲线

Fig. 1 Hilbert curve

1.4 位置本地差分隐私

存在 n 个位置，每个位置对应一条记录。给定一个隐私算法 N 及其定义域 $Def(N)$和值域 $Ran(N)$，若任意两条位置记录 t 和 t'($t,t'\in Def(N)$)都满足相同的输出结果 t^*($t^*\in Ran(N)$)，且满足以下不等式，则算法 N 满足 ε-本地化差分隐私：

$$P(N(t)=t^*)\le e^{\varepsilon}P(N(t')=t^*) \tag{4}$$

由此可知，本地差分隐私通过控制算法 N 输出相似的结果，使攻击者无法区分哪一条数据为用户的真实数据。

2 位置隐私保护方案

2.1 系统结构

在基于可信第三方的模型中，用户发起多次请求时，TTP 容易变成影响系统效率的阻碍。并且，TTP 本身容易受到攻击，一旦 TTP 被攻破，用户的全部隐私信息都会泄露。因此本文所采用的方案不使用 TTP，主要由本地用户和 LSP 两部分构成，如图 2 所示。本地用户通过无线设备获取自身位置信息，并在本地运行隐私保护方案，避免使用不可信第三方造成的安全隐患。将扰动后的位置查询发送到 LSP，由 LSP 提供查询结果返回给本地用户，通过位置处理算法对查询结果进行处理，将用户所需的数据展现给用户。

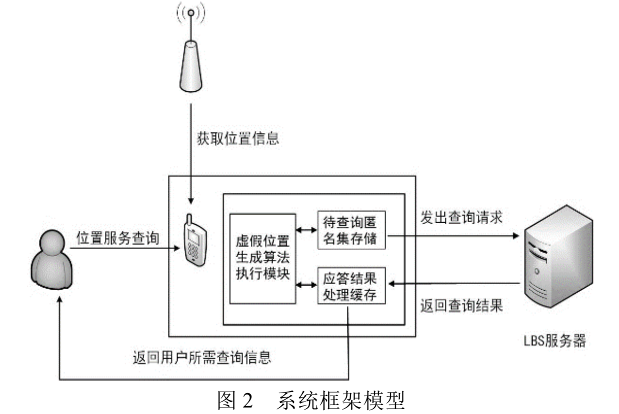


图 2 系统框架模型

Fig. 2 System framework model

2.2 k-匿名位置集生成

在 k-匿名集生成过程中，首先根据用户历史查询记录获

取用户在城市兴趣点中提交查询请求的概率，将兴趣点的概率按大小排序生成概率表 T，将 T 存入本地便于用户之后的查询，对 T 定期更新防止其数据落后。针对攻击者利用背景知识进行攻击，应当确保匿名集的熵尽可能的大，因此将与用户真实位置 Z 查询概率最接近的点添加到匿名候选区 L_c 中。由文献[13]可知，L_c 中的熵随着位置数的增加而增多，当位置数达到 $2k$-2 时接近最大熵，且不再明显增大。位置数的数量直接影响了计算开销，为了权衡效率和隐私保护性，将 L_c 中的位置数量设为 $2k$-2。为了保证匿名集中所选位置具有较好的效用，在 L_c 生成过程中，以用户真实位置为中心，根据欧几里德距离选择最近的兴趣点。生成 L_c 的算法如下：

算法 1 匿名候选区 L_c 生成算法

输入：T，Z，k

输出：L_c

a) 从 T 中获取 Z 的查询概率 Z_p

b) 获取与 Z_p 差值不超过 ρ 的兴趣点，存入临时位置集 R 中

c) 根据 $dis(r_i,Z)=\sqrt{(x_r-x_Z)^2+(y_r-y_Z)^2}$ 计算 R 中各个位置 r_i 到 Z 的欧几里德距离 S_i

d) 堆排序法取 S_i 最小的前 $2k$-2 个兴趣点，存入 L_c

e) 结束

如图 3 所示，通过查询表 T 获取地图中兴趣点概率后与 Z 比较，可得兴趣点 L_1，L_2，L_3，L_4，L_5 与 Z 概率差值不超过 ρ=0.01，计算选中的兴趣点与 Z 的欧几里德距离 $S_{i=\{1,2,3,4,5\}}$。设 k=2，比较 S_i 后得 L_c={L_2，L_3}。设临时位置集 R 中含有 n 个位置点，算法 1 的空间复杂度为 $O(n)$。使用堆排序对 R 中位置点到 Z 的距离进行排序，时间复杂度为 $O(n\log n)$。

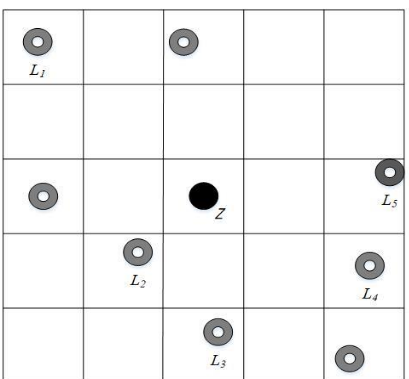


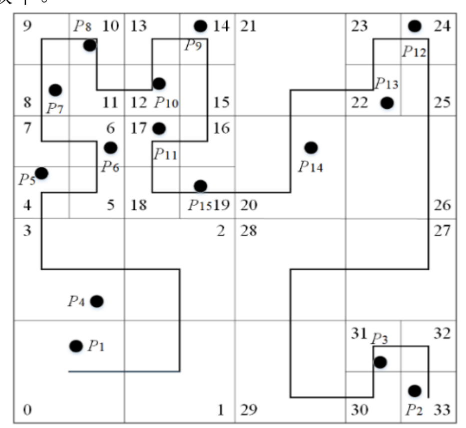
图 3 候选区兴趣点生成

Fig. 3 Candidate area interest point generation

2.3 IHC 划分兴趣点

在生成的 L_c 中随机取 k-1 个位置与用户真实位置组成 k-匿名集 L。在生成 L 的过程中，应当尽可能的使所取位置点分散，保证匿名区域的范围，将各个兴趣点之间的欧几里德距离之和作为衡量离散度的标准，通过多次随机分配，选取位置点最分散的匿名集。这时攻击者从 L 中得到用户真实位置的概率接近 $1/k$。由于本文选择匿名位置的标准之一是与真实位置的欧几里德距离，所产生的匿名位置会以真实位置为中心，这不利于真实位置的保护。为了进一步降低攻击者获取真实位置的概率，使用改进的 Hilbert 曲线(IHC, Improved Hilbert curve)对位置进行划分，将划分后的地图通过 RAPPOR 进行扰动。

Hilbert 曲线能够将地理位置从二维空间映射到一维空间上，能够保证在空间上相邻的点投射到一维空间后也相邻，减少了数据处理时间，提高了数据处理效率。然而 Hilbert 曲线无法反映兴趣点的密集度分布，通常兴趣点密集的位置应当使用更细的粒度划分。IHC 的构建如下：将 L 中距离 Z 最远的点作为边界 R，使所有位置点包含在 $N\times N$ 的地图空间中。当区域内的兴趣点数大于阈值 σ=1 时，将该区域递归的划分成 4 个大小相同的正方形子区域。图 4(a)为位置点划分后的区域分布，保证了位置点的密度分布。将划分后的 IHC 存储到四叉树中，存储方式如图 4(b)，兴趣点数量为 k，文件的存储开销为 $O(k)$，计算各个兴趣点 IHC 值的时间复杂度为 $O(k)$。与 Hilbert 曲线相比，IHC 划分可以节省存储空间，提升计算效率。



(a) IHC division

(b) IHC division area representation and storage

图 4 IHC 映射

Fig. 4 IHC mapping

2.4 基于本地差分隐私的 RAPPOR 扰动

RAPPOR[22]能够将终端用户众包数据进行匿名，提供了高效的隐私和效用，且不依赖可信第三方。对于映射到 Hilbert 曲线上的匿名候选位置，通过 RAPPOR 可实现随机扰动，获得强大的隐私保护。设 A={a_1，a_2，…，a_n}为地图划分后的区域 ID，n 为划分后的区域总数。对于第 i 个区域，如果存在选中的兴趣点，a_i 设置为 1；否则 a_i 设置为 0。设 R 为 n 位数组，R_j 表示 R 中第 j 位的值，当 a_j 为 1 时，R 对应的位设为 1，其他位设为 0，如公式(5)所示。

$$R_j=\begin{cases}1, & \text{if } a_j=1\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

接下来是扰动从上式中获取的 R。R 中每个位进行随机响应扰动，如公式(6)所示。

$$P(R_j'=x)=\begin{cases}0.5f, & x=1\\ 0.5f, & x=0\\ 1-f, & x=R_j\end{cases} \tag{6}$$

其中，f($f\in$[0，1])为控制隐私级别的概率参数，越接近 1 的值拥有更强的隐私保证。在 RAPPOR 中，产生的 R'被称为永久随机响应。

然后，将另一个扰动施加到 R'的每一位，得到瞬时随机响应，表示为 U，如公式(7)所示。

$$P(U_j=1)=\begin{cases}q, & \text{if } R_j'=1\\ p, & \text{if } R_j'=0\end{cases} \tag{7}$$

生成的 U 在 RAPPOR 中称为瞬时随机响应，其第 k 位设置为 1 的概率受参数 q(或 p)和 R_k 影响。根据 RAPPOR，上述随机编码方法满足 ε-差分隐私。

初始被选中区域经过扰动后仍被选中的概率为

$$q^*=P(U_i=1\mid R_i=1)=\frac{1}{2}f(p+q)+(1-f)q \tag{8}$$

初始未被选中区域扰动后被选中的概率为

$$p^*=P(U_i=1|R_i=0)=\frac{1}{2}f(p+q)+(1-f)p \quad (9)$$

$$\varepsilon=k\ln\left(\frac{q^*(1-p^*)}{p^*(1-q^*)}\right) \quad (10)$$

RAPPOR 扰动后，通过 IHC 解码获取地理位置集 R，使用 R 进行查询即可保证用户位置隐私。k-匿名集经过扰动后，用户真实位置可能在扰动中丢失，当真实位置不在 R 中时，对于 LBS 服务器返回的匿名集的查询结果，获取与用户真实位置欧几里德距离最近的 n 个位置点的查询结果，对所获取的位置信息取并集，作为用户的查询信息。位置并集算法如算法 2 所示。

算法 2　位置并集获取算法

输入：$R=\{l_t, t=1, 2, \cdots, r\}$

输出：结果集 T

a) 设 T 为空

b) 对于 R 中所有位置点，计算与用户真实位置的欧几里德距离

c) 堆排序取与用户位置距离最小的前 n 个位置

d) 选取 1 个位置，将其查询结果的兴趣点存入 T

e) 依次查询 n 个位置的查询结果，与 T 求并集后存入 T

f) 返回 T

图 5 为位置集 R 中兴趣点的查询结果。位置集 $R=\{L_1, L_2, L_3, L_4, L_5\}$为扰动后选取到的五个兴趣点，$n$ 取 3，选取欧几里德距离最小的前三个位置点为 L_2，L_4，L_5，其查询结果为 $L_2=\{a, b, e, f\}$，$L_4=\{e, f, g\}$，$L_5=\{c, d\}$，则取并集后用户获得的查询信息为 $T=\{a, b, c, d, e, f, g\}$。设 R 中含有 r 个位置点，则堆排序的时间复杂度为 $O(r\log r)$，设每个位置含有 m 个查询结果，生成查询结果 T 的时间复杂度为 $O(nm\log n)$,因此算法 2 的时间复杂度为 $max(O(r\log r), O(nm\log n))$。

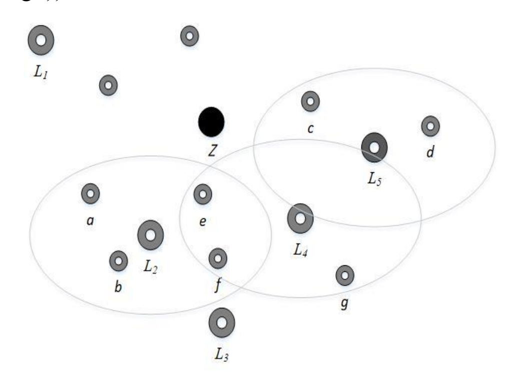


图 5　位置集查询结果

Fig. 5　Location set query results

3　实验与性能评估

3.1　实验环境与方法

采用旧金山数据集[25]验证所设计方案的性能。该数据集包含 174956 个兴趣点。如图 6 所示，x 和 y 分别表示经纬度转换成的直角坐标。使用 Python3.6 编程实现，操作系统为 Windows 10 家庭版，计算机 CPU 型号为 Intel i7，内存容量为 64GB。

通过 LBS 服务器可以获取用户历史查询记录。由文献[26]可知，当无法获取查询记录时，可以通过地图上的兴趣点数代替用户查询记录。本实验使用旧金山数据集的兴趣点作为用户查询记录。将地图分割为 100×100 相同大小的位置单元，计算每个位置单元历史查询概率作为先验概率，在每个位置单元中随机选择一个兴趣点。兴趣点的查询概率为相应位置单元的历史查询概率，用户所在位置单元的兴趣点为其真实位置点。

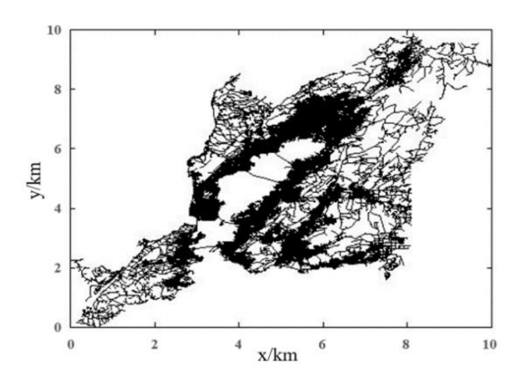


图 6　旧金山数据集

Fig. 6　San francisco dataset

3.2　安全性分析

从匿名集的熵和攻击算法识别概率方面进行安全性分析，并将本文的方案与其他方案进行比较。其中每组数据取 100 次实验的平均值，k 的取值范围为 2~30。

3.2.1 攻击方法分析

对于 LBS 服务，主要有背景攻击、概率攻击和语义攻击。对于背景攻击，位置隐私通常通过消除背景信息和用户当前位置之间的联系来保护。

概率攻击指攻击者通过已知信息筛选出不合理的位置点如河流、沙漠等，从而提高发现用户真实位置的概率。这些位置点与用户真实位置没有直接关联，但是通过过滤 k-匿名集中一些虚假位置点，使匿名集不满足 k-匿名要求，降低隐私保护水平，起到了辅助攻击的效果。通常情况下，对于概率攻击，可以在获取查询用户的历史查询记录后，将查询概率高的位置作为虚假位置点，迷惑攻击者。本文将真实兴趣点作为虚假位置，选择与用户真实位置查询概率相同的位置点构成候选区集合，有效避免了概率攻击的发生。

语义攻击的形式很多，其中位置同质攻击是语义攻击中一种常见的攻击手段。位置同质攻击是当匿名位置和用户真实位置间的距离过于接近时，即使达到了 k-匿名要求，但是匿名区域太小，攻击者可以通过位置聚类等方法进一步缩小匿名区域，从而增大获取用户真实位置的概率。本文方案选择位置分散度最大的位置集合作为匿名集，降低了攻击者利用位置同质攻击获取真实位置的概率。

3.2.2 位置熵

为了更好验证本文方案的性能，使用文献[13] [14]和最优选择进行对比。由式 3 可知，k 不变时位置熵的大小受匿名集中兴趣点的查询概率影响，查询概率之间差值越小，熵越大，当匿名集中所有位置查询概率都相等时，所生成匿名集的位置熵最大，本文将这种情况定义为最优选择。每个方案实验结果如图 7 所示。

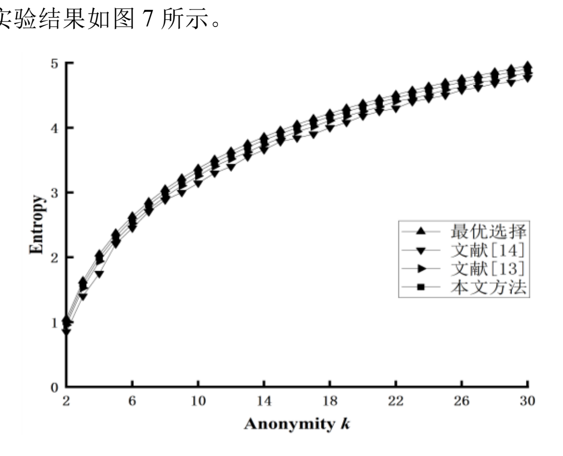


图 7　位置熵值

Fig. 7　Location entropy

由图 7 可以看到，生成 *k*-匿名集的位置熵随着 *k* 的增大而增大，即匿名集的安全性增强。其中最优选择法的熵值最大，文献[13]和文献[14]在生成匿名集的过程中充分考虑了兴趣点的查询概率，选取与用户位置查询概率相近的兴趣点构成匿名集，因此所选位置的熵值与最优选择相近，分别比最优选择平均低 0.5%和 3%。因为本文方法在设计中考虑了攻击者可知的背景信息，尽可能的保证了 *k*-匿名集中的位置点查询概率相同。本文在算法 1 中引入了差值参数 ρ，通过减小 ρ 选取概率最近的兴趣点构造匿名集，生成的匿名集的熵值与最优选择法最接近，仅比其平均低 0.3%，因此本文方法所生成的匿名集具有很高的隐私保护度。

3.2.3 攻击算法识别用户位置概率

当攻击者获取用户所发送匿名集 *R* 后，结合背景信息对用户位置发起攻击，通过[14]中所用攻击算法推断用户的真实位置。攻击者获取用户发送位置集后，结合边信息推断用户位置分布概率，确定用户真实位置。图 8 为攻击算法推断文献[13]、[14]、[17]和本文方案在不同隐私度 *k* 下匿名集位置分布的概率。式 6 中 *f* 的取值决定初始扰动的程度，*f* 取 1 时为完全随机响应，取 0 时为无扰动，为了保证隐私和效用的平衡，取 *f* 为 0.5。式 7 中 *q*，*p* 决定查询位置集的扰动程度以及位置集中的兴趣点数量，*p*+*q* 值越大，查询集中位置点越多，*p*+*q*=1 可保证扰动前后兴趣点数目基本不变，通过分析查询结果的效用，*q*，*p* 分别取 0.75，0.25。通过实验得算法 2 中 *n* 取值为 *k*/2 时，生成的查询结果与真实位置查询结果基本相同，且效率最高，因此本文中 *n* 取值为 *k*/2。

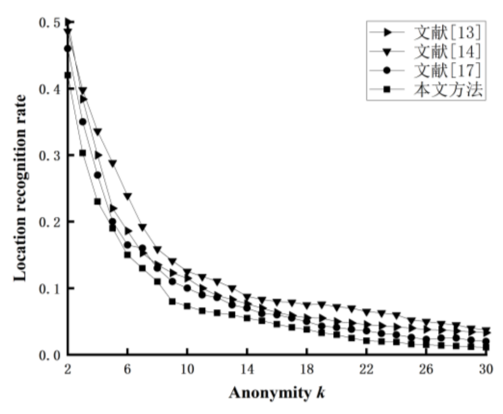


图 8　位置识别概率

Fig. 8　Probability of Location Recognition

由图 8 可知，本文方案所生成的匿名集中用户真实位置被识别的概率要低于其他三种方案，相比于文献[13][14],本方案添加了 RAPPOR 扰动，使用户真实位置出现不存在于匿名集的情况，增加了匿名集的随机性。文献[17]选取混淆位置时未考虑背景知识信息，通过背景知识可缩小用户位置范围。本方案匿名区域随着 *k* 值的增加而增大，降低了攻击者通过语义攻击等方法获取用户位置的概率，有效提升了用户位置保护效果。

3.3　性能分析

算法在保证用户位置隐私安全性的同时应当充分考虑其性能效用，图 9 表示文献[13][14][17]和本文方案查询位置服务可用性比较。由图 9 可以看出，随着 *k* 值增大，服务可用性降低，本文方法生成匿名集的同时能够更好的保证查询结果的可用性，这是因为本文的匿名集中存在包含用户真实位置和不包含真实位置两种情况。当包含真实位置时，位置可用性为最优；当不包含用户真实位置时，通过算法 2 获取用户真实位置附近的查询结果集，可以保证查询结果的可用性。相同情况下本文方法比文献[13][14][17]分别平均高 11.95%、5.92%和 29.51%。

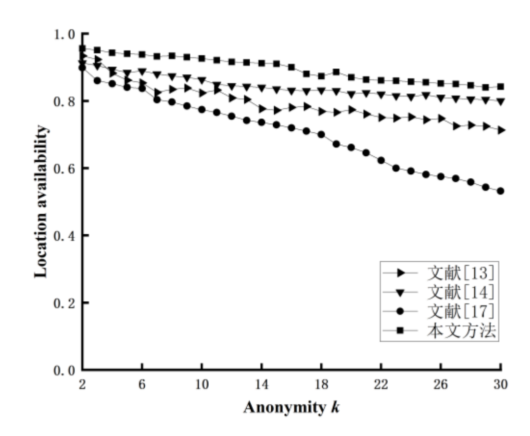


图 9　位置服务可用性

Fig. 9　Location service availability

图 10 为文献[13][14][17]和本文方案的时间开销。通过对比可得知，随着 *k* 值的增加，算法执行时间逐渐增多。文献[14]构造匿名集的同时需要对位置进行离散选择，因此其运行时间比其他方法略长。本文方法由于添加了扰动，导致其时间略高于文献[13]。

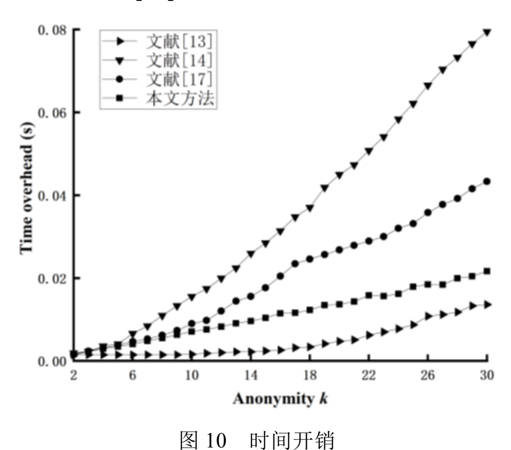


图 10　时间开销

Fig. 10　Time overhead

4　结束语

对于位置服务中存在的用户位置隐私保护的问题，本文对现有位置隐私保护方法进行研究，提出基于本地差分隐私的匿名保护方案，对 *k*-匿名集中的候选位置进行扰动，降低了用户真实位置泄露的概率。通过安全性和可用性分析，证明了本方案对于用户隐私和效用起到了很好的权衡，性能得到了明显提升。

本文方案在使用匿名集进行查询访问时，存在查询结果利用率低的问题。在接下来的工作中，本文将通过使用本地缓存的方式，提高查询资源利用率，降低用户和 LBS 服务器交互的次数，提升用户位置隐私保护程度。

参考文献：

[1] Dilay P, Udai P R. Towards Privacy-Preserving Dummy Generation in Location-Based Services [J]. Procedia Computer Science, 2020, 2020 (171): 1323-1326.

[2] Seo Y D, Cho Y S. Point of interest recommendations based on the anchoring effect in location-based social network services [J]. Expert Systems with Applications, 2021, 2021 (164): Article ID 114018.

[3] 张青云, 张兴, 李万杰, 等. 基于 LBS 系统的位置轨迹隐私保护技术综述 [J]. 计算机应用研究, 2020, 37 (12): 3534-3544. (Zhang Qingyun, Zhang Xing, Li Wanjie, *et al.* Overview of location trajectory privacy protection technology based on LBS system [J]. Application Research of Computers, 2020, 37 (12): 3534-3544.)

[4] Sweeney L. k-Anonymity: A Model for Protecting Privacy [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10 (05): 557-570.

[5] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking [C]// Proc of the 1st International Conference on Mobile Systems, Applications and Services. New York: ACM Press, 2003: 31-42.

[6] Kido H, Yanagisawa Y, Satoh T. An anonymous communication technique using dummies for location-based services [C]// ICPS'05. Proceedings. International Conference on Pervasive Services, 2005. Piscataway, NJ: IEEE Press, 2005: 88-97.

[7] Zhu Xiaoyan, Chi Haotian, Niu Ben, *et al.* Mobicache: When k-anonymity meets cache [C]// 2013 IEEE Global Communications Conference (GLOBECOM) . Piscataway, NJ: IEEE Press, 2013: 820-825.

[8] 叶阿勇，李亚成，马建峰，等．基于服务相似性的 k-匿名位置隐私保护方法 [J]. 通信学报, 2014, 35 (11): 162-169. (Ye Ayong, Li Yacheng, Ma Jianfeng. *et al.* K-anonymous location privacy protection method based on service similarity [J]. Journal on Communications, 2014, 35 (11): 162-169.)

[9] Yin Chunyong, Xi Jinwen, Sun Ruxia. Location Privacy Protection Based on Improved K-Value Method in Augmented Reality on Mobile Devices [J]. Mobile Information Systems, 2017, 2017 (12): 1-7.

[10] Jin Lei, Li Chao, Palanisamy B, *et al.* k-Trustee: Location injection attack-resilient anonymization for location privacy [J]. Computers & Security, 2018, 78 (2): 212-230.

[11] Ling Jie, Xu Junyi. Decentralized Location Privacy Protection Method of Offset Grid [C]// 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019) . Atlantis Press, 2019: 113-120.

[12] Zhang Yongbing, Zhang Qiuyu, Yan Yan, *et al.* A k-Anonymous Location Privacy Protection Method of Polygon Based on Density Distribution [J]. International Journal of Network Security, 2021, 23 (1): 57-66.

[13] 闫光辉，刘婷，张学军，等．抵御背景知识推理攻击的服务相似性位置 k 匿名隐私保护方法 [J]. 西安交通大学学报, 2020, 54 (01): 8-18. (Yan Guanghui, Liu Ting, Zhang Xuejun, *et al.* Service similarity location k anonymous privacy protection method resisting background knowledge reasoning attack [J]. Journal of Xi'an Jiaotong University, 2020, 54 (01): 8-18.)

[14] 杨洋，胡晓辉，杜永文．基于历史查询概率的 K-匿名哑元选取算法 [J/OL]. 计算机工程：1-14. (2021-03-30) [2021-12-10] https://doi. org/10. 19678/j. issn. 1000-3428. 0060417. (Yang Yang, Hu Xiaohui, Du Yongwen. K-anonymous dummy selection algorithm based on historical query probability [J/OL]. Computer Engineering: 1-14. (2021-03-30) [2021-12-10] https://doi. org/10. 19678/j. issn. 1000-3428. 0060417.)

[15] DWORK C. Differential privacy [C]// International Colloquium on Automata, Languages, and Programming. Berlin: Springer, 2006: 1-12.

[16] 袁健，王迪，高喜龙，等．基于差分隐私的匿名组 LBS 轨迹隐私保护模型 [J]. 小型微型计算机系统, 2019, 40 (02): 341-347. (Yuan Jian, Wang Di, Gao Xilong, *et al.* An anonymous group LBS trajectory privacy protection model based on differential privacy [J]. Journal of Chinese Computer Systems, 2019, 40 (02): 341-347.)

[17] Wang Jie, Wang Feng, Li Hongtao. Differential Privacy Location Protection Scheme Based on Hilbert Curve [J]. Security and Communication Networks, 2021, 2021 (1): Article ID 5574415.

[18] Zhang Yongbing, Zhang Qiuyu, Li Zongyi, *et al.* A k-anonymous Location Privacy Protection Method of Dummy Based on Geographical Semantics [J]. International Journal of Network Security, 2019, 21 (6): 937-946.

[19] Zhang Qingyun, Zhang Xing, Wang Mingyue, *et al.* DPLQ: Location - based service privacy protection scheme based on differential privacy [J]. IET Information Security, 2021, 15 (6): 442-456.

[20] DUCHI J C, JORDAN M I, WAINWRIGHT M J. Local privacy and statistical minimax rates [C]// 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. Piscataway, NJ: IEEE Press, 2013: 429-438.

[21] FANTI G, PIHUR V, ERLINGSSON Ú. Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries [J]. Proc on Privacy Enhancing Technologies, 2016, 2016 (3): 41-61.

[22] ERLINGSSON Ú, PIHUR V, KOROLOVA A. Rappor: Randomize aggregatable privacy-preserving ordinal response [C]// Proc of the 2014 ACM SIGSAC conference on computer and communications security. New York: ACM Press, 2014: 1054-1067.

[23] Wang Xiongjian, Yang Weidong. Protection method of continuous location uploading based on local differential privacy [C]// 2020 International Conference on Networking and Network Applications (NaNA) . Piscataway, NJ: IEEE Press, 2020: 157-161.

[24] Wang Jian, Wang Yanli, Zhao Guosheng, *et al.* Location protection method for mobile crowd sensing based on local differential privacy preference [J]. Peer-to-Peer Networking and Applications, 2019, 12 (5): 1097-1109.

[25] BRINKHOFF T. A framework for generating network-based moving objects [J]. GeoInformatica, 2002, 6 (2): 153-180.

[26] PINGLEY A, Zhang Nan, Fu Xinwen, *et al.* Protection of query privacy for continuous location based services [C]// Proceedings of the Computer and Communications Societies. Piscataway, NJ: IEEE Press, 2011: 1710-1718.